Development of a Multi-Fidelity Reduced-Order Model Based on Manifold Alignment

Christian Perron *, Dushhyanth Rajaram †, and Dimitri N. Mavris ‡
Georgia Institute of Technology, Atlanta, Georgia, 30332

This work presents the development of a novel multi-fidelity, parametric, and non-intrusive Reduced Order Modeling (ROM) method to tackle the problem of achieving an acceptable predictive accuracy under a limited computational budget, i.e., with expensive simulations and sparse training data. Traditional multi-fidelity surrogate models that predict scalar quantities address this issue by leveraging auxiliary data generated by a computationally cheaper lower fidelity code. However, for the prediction of field quantities, simulations of different fidelities may produce high-dimensional responses with inconsistent dimensionality and topology, rendering the direct application of common multi-fidelity techniques challenging. The proposed approach uses manifold alignment to fuse inconsistent fields from high- and low-fidelity simulations by individually projecting their solution onto a common shared latent space. Hence, simulations using incompatible grids or geometries can be combined into a single multi-fidelity ROM without additional manipulation of the data. This method is applied to a variety of multi-fidelity scenarios using a transonic airfoil problem. In most cases, the new multi-fidelity ROM achieves comparable predictive accuracy at a substantially lower computational cost. Furthermore, it is demonstrated that the proposed method can readily combine disparate fields without any adverse effect on model performance.

Nomenclature

$\mathbf{X}, \mathbf{Y}$	=	high- and low-fidelity training samples
$\mathbf{Z}, \mathbf{W}$	=	high- and low-fidelity latent coordinates
$\mathbf{Z}_{\text{lf}}$	=	low-fidelity approximation of $\mathbf{Z}$
$\mathbf{P}$	=	design parameters
$\mathbf{U}, \mathbf{V}$	=	left and right singular vectors
$\mathbf{Q}$	=	rotation matrix
$\mathbf{S}$	=	sample covariance matrix
$\mathbf{\Phi}_k, \mathbf{\Psi}_k$	=	high- and low-fidelity POD modes
$\mathbf{\Sigma}$	=	diagonal matrix of singular values
Re	=	Reynolds number
M	=	Mach number
C_n, C_d, C_m	=	normal, drag, and pitching moment coefficients
C_P	=	coefficient of pressure
f_x, f_y	=	high- and low-fidelity model
g	=	regression model
d, q	=	high- and low-fidelity field dimensionality
k	=	latent space dimensionality
b	=	number of design parameters
n, m	=	number of high- and low-fidelity training samples
n_{test}	=	number of test samples
s	=	scaling factor
α	=	angle of attack
λ	=	eigenvalue of $\mathbf{S}$

*Senior Graduate Researcher, ASDL, Daniel Guggenheim School of Aerospace Engineering, AIAA Student Member
†Senior Graduate Researcher, ASDL, Daniel Guggenheim School of Aerospace Engineering, AIAA Student Member
‡S.P. Langley Distinguished Regents Professor and Director of ASDL, Daniel Guggenheim School of Aerospace Engineering, AIAA Fellow

τ = multi-fidelity ratio
ε = POD reconstruction residual
$\epsilon_p, \epsilon_{rc}, \epsilon_{rg}$ = prediction, reconstruction, and regression errors

Subscript
i = row index referring to i-th coordinate
j = column index referring to j-th sample
L, U = linked and unlinked data
∞ = free-stream value

Superscript
$-$ = sample average
$\sim$ = model prediction
$*$ = out-of-sample value

I. Introduction and Background

As the aerospace industry continuously strives towards more efficient and cleaner aircraft, aerospace engineers are compelled to use higher fidelity simulations earlier in the design process. Yet, despite the increased availability of high-performance supercomputers, high-fidelity simulations such as Computational Fluid Dynamics (CFD) still require many hours, if not days, to solve. As such, the use of these expensive models together with design space exploration, numerical optimization, or uncertainty quantification, can easily incur a prohibitive computational cost.

A. Reduced-Order Modeling

In recent years, projection-based Reduced-Order Modeling (ROM) [1] has emerged as a powerful tool capable of reconciling the desire for high-fidelity results with the need for computationally inexpensive models. Beginning with a dataset of previously generated solutions, this class of methods addresses the computational complexity of high-fidelity models by extracting a low-dimensional subspace or latent space that best retains the fundamental features of the physical problem [2]. The high-dimensional solution of a given model is thus reduced to the linear combination of a handful of dominant modes. A popular method for extracting this low-dimensional subspace is the Proper Orthogonal Decomposition (POD). Based on the work by Lumley [3], it is closely related to the Karhunen-Loève Transformation (KLT) and the Principal Component Analysis (PCA) [4]. Some ROM methods project the governing equation of the physical model onto the extracted subspace using the Galerkin method [1]. This category of ROMs is regarded as intrusive and has the disadvantage of requiring access to the underlying simulation source code. Alternatively, one of the most popular non-intrusive ROM methods combines POD with a regression method in the latent space to predict new solutions from previously generated results [5–9], essentially treating the original model as a black-box. In this sense, non-intrusive ROMs are similar to traditional surrogate models, but instead of estimating some integrated quantities (e.g., lift, drag, moment), they predict high-dimensional field solutions of a simulation model, thus providing a physically richer result. Although generally less accurate than intrusive methods for highly non-linear problems [10], non-intrusive methods are more versatile and their implementation is independent of the underlying physical model. In this sense, they are relatively more practical and flexible than intrusive methods. Therefore, this work focuses on non-intrusive ROMs.

B. Multi-Fidelity ROM

A notable drawback of non-intrusive ROMs is that the accuracy of their prediction relies heavily on the quality of the underlying regression model. For highly non-linear responses with many design parameters, most regression methods require a large amount of training data to obtain a reasonable accuracy. In the context of expensive high-fidelity simulations, this usually represents a substantial computational cost. With traditional surrogate models, aircraft designers often rely on multi-fidelity methods to alleviate the high training cost by leveraging a cheaper low-fidelity model [11–14]. In principle, multi-fidelity surrogate models combine sparse high-fidelity and dense low-fidelity datasets to achieve similar prediction accuracy as their single-fidelity counterpart albeit at a lower training cost. While multi-fidelity methods abound for traditional surrogate models, few exist for reduced-order modeling. One major difficulty with multi-fidelity ROMs is the potential inconsistency existing between solutions produced by analyses of varying fidelities.

For instance, two simulations can use grids of different sizes which would create fields of inconsistent dimensionality. The topology of both solutions can also be incompatible in cases where a simplified geometry is used for the low-fidelity model. As an example, in aerodynamic analysis, a Vortex Lattice Method (VLM) represents an aircraft using thin surfaces while a CFD simulation uses the actual 3D geometry as illustrated in Fig. 1.

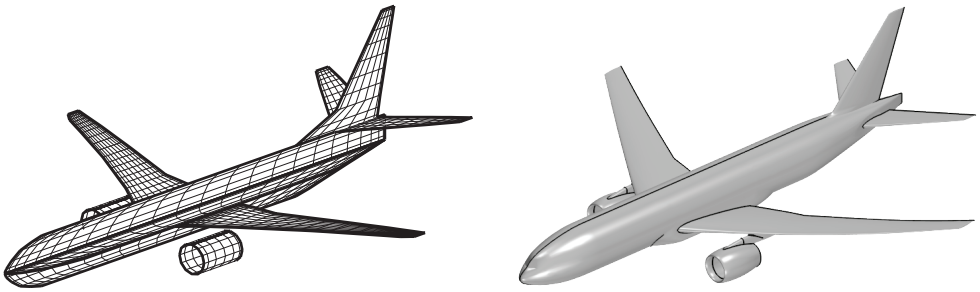
To circumvent any potential inconsistency between the high- and low-fidelity solutions, some authors have proposed multi-fidelity ROMs that use the same grid and geometry for both fidelity levels but rely on a simpler numerical method for the lower-fidelity simulation to achieve some computational savings. This approach was adopted by Mifsud et al. [15, 16] for the prediction of the pressure distribution over a projectile and by Bertram et al. [17] for the aerodynamic design of a road vehicle. Although using a common geometry and discretization avoids any compatibility issues between simulations, it severely restricts the benefits of a multi-fidelity method since great computational savings can be obtained by using a coarser grid.

An alternative solution is to map the results of all fidelity levels onto a common grid during a pre-processing step, i.e., before building the ROM. This strategy was used by Benamara et al. [18, 19] who performed CFD simulations of a compressor blade using a coarse and fine grid, then interpolated all the results onto the coarse grid. In their work, they extracted an initial POD basis using the high-fidelity solution and then extended this subspace using basis vectors obtained from the low-fidelity solutions. Similarly, Malouin et al. [20] computed the difference between the coarse and fine solutions of the flow around an airfoil, then created a ROM of the resulting discrepancy field. A drawback of interpolating all the solutions onto a common grid is the potential introduction of interpolation errors into the results. Besides, mapping results from a fine grid onto a coarse grid, as done by both Benamara et al. and Malouin et al., degrades the resolution of the high-fidelity simulations. Furthermore, this mapping strategy would not be possible if the different fidelity levels use incompatible geometries such as depicted in Fig. 1.

Another multi-fidelity ROM proposed by Toal [21] and later used by Benamara et al. [22] is based on Gappy-POD, a method originally developed by Everson and Sirovich [23] for the reconstruction of incomplete data [5, 24]. In this approach, the ROM prediction of a new solution is achieved by reconstructing a high-fidelity result from a new low-fidelity solution, essentially using the low-fidelity result as an input to the ROM. In some sense, the Gappy-POD method converts the multi-fidelity problem into a missing value problem. While this multi-fidelity ROM can be applied to any combination of models without requiring the low- and high-fidelity results to be consistent, a new low-fidelity solution is required to make a high-fidelity prediction. As such, the evaluation speed of this ROM is restricted by the execution time of the low-fidelity simulation.

C. Contributions of Current Work

In an attempt to solve the limitations of the previous multi-fidelity methods, this paper develops a novel parametric, non-intrusive, and multi-fidelity reduced-order modeling method that leverages manifold alignment [25, 26]. The proposed method resolves the issue of inconsistent fields by individually projecting both the high- and low- fidelity results onto a common latent subspace using manifold alignment. As shown later in this paper, manifold alignment is applicable to disparate high- and low-fidelity results under the mild and often true assumption that both datasets are correlated. Furthermore, the method is non-intrusive and unlike the Gappy-POD based approach of Toal [21], new predictions do not require additional low-fidelity results. The following paper is divided as follows. Section II reviews



a) Thin surfaces b) Actual geometry

Fig. 1 Comparison of a thin surface representation used for VLM simulations and a 3D geometry used for CFD simulations.

the conventional POD-based ROM and introduces the new multi-fidelity method based on manifold alignment. To assess the performance of the proposed method, details of a 2D transonic airfoil test case are introduced in Section III. Finally, Section IV evaluates the results of the experiments and compares the performance of the proposed multi-fidelity ROM to an equivalent single-fidelity POD-based approach.

II. Numerical Method

This section provides details of the numerical methods used in this work. A brief overview of the standard POD technique is followed by a description of manifold alignment, which is employed to merge fields of different fidelities. Then, noteworthy comments regarding the interpolation of latent coordinates using multi-fidelity surrogate models are presented. Finally, we demonstrate how all the presented methods are integrated into a multi-fidelity, parametric, and non-intrusive ROM.

A. Proper Orthogonal Decomposition

The POD is a linear and unsupervised dimensionality reduction technique that is employed extensively in projection-based ROMs. For all practical purposes, POD is equivalent to PCA, the latter name being preferred in the fields of statistical analysis and machine learning. The following section provides a brief overview of the POD procedure.

Consider the matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ containing fields obtained for a range of design parameter values, whose j-th column corresponds to the sample $\mathbf{x}_j \in \mathbb{R}^d$. The samples $\mathbf{x}_j$ are often referred to as *snapshots* in the literature [27]. For convenience and without loss of generality, we assume that the results in $\mathbf{X}$ have been centered beforehand such that the sample mean vector $\overline{\mathbf{x}} \in \mathbb{R}^d$ is

$$\overline{\mathbf{x}} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j = \mathbf{0} \tag{1}$$

The objective of the POD is to find a set of orthonormal vectors $\boldsymbol{\phi}_j \in \mathbb{R}^d$, often called POD modes, along which the variance of the data is maximum. In mathematical terms, the solution of POD is given by the optimization of the following Rayleigh quotient [28]

$$\max_{\boldsymbol{\phi}_j} \frac{\boldsymbol{\phi}_j^T \mathbf{S} \, \boldsymbol{\phi}_j}{\boldsymbol{\phi}_j^T \boldsymbol{\phi}_j} \tag{2}$$

where $\mathbf{S} \in \mathbb{R}^{d \times d}$ is the sample covariance matrix defined as

$$\mathbf{S} = \frac{1}{n} \mathbf{X} \mathbf{X}^T \tag{3}$$

The solution to (2) is given by the following eigenproblem

$$\mathbf{S} \boldsymbol{\phi}_j = \lambda_j \boldsymbol{\phi}_j \quad \text{s.t.} \quad \|\boldsymbol{\phi}_j\| = 1 \tag{4}$$

where λ_j is the j-th eigenvalue of $\mathbf{S}$ in decreasing order and $\boldsymbol{\phi}_j$ is the associated eigenvector. The objective of Eq. (2) is maximized when $\boldsymbol{\phi}_j$ is the eigenvector corresponding to the largest λ_j. Hence, we define the POD latent space using the orthonormal basis $\boldsymbol{\Phi}_k \in \mathbb{R}^{d \times k}$ whose columns are composed of the eigenvectors with the k largest eigenvalues, where $k \ll d$. The dimensionality k of the POD latent space can be chosen using the Relative Information Content (RIC) [29, 30] criterion defined as

$$\text{RIC} = \frac{\sum_{j=1}^{k} \lambda_j}{\sum_{j=1}^{d} \lambda_j} \tag{5}$$

The RIC essentially represents the fraction of the total variance captured by the POD basis. For reduced-order modeling applications, RIC values of 99.9% or higher are fairly common.

The projection of a sample $\mathbf{x}_j$ onto this reduced subspace is given by $\mathbf{z}_j = \boldsymbol{\Phi}_k^T \mathbf{x}_j$, where $\mathbf{z}_j \in \mathbb{R}^k$ is a latent variable representing the coordinates of the j-th sample in the POD latent space. Conversely, the approximate reconstruction of the field $\widetilde{\mathbf{x}}_j$ is then obtained by $\widetilde{\mathbf{x}}_j = \boldsymbol{\Phi}_k \mathbf{z}_j$. If we consider the reconstruction residual ε defined as

$$\varepsilon = \frac{1}{n} \sum_{j=1}^{n} \|\mathbf{x}_j - \widetilde{\mathbf{x}}_j\|^2 \tag{6}$$

it can be shown that the POD basis defines the linear subspace that provides the optimal reconstruction residual in a least-squares sense [31]. This last fact is one of the reasons behind the effectiveness of POD for reduced-order modeling.

B. Manifold Alignment

In a multi-fidelity context, let us consider the datasets $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{d\times n}$ and $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_m] \in \mathbb{R}^{q\times m}$ with columns holding field solutions of a high- and low-fidelity model respectively. Even though $\mathbf{X}$ and $\mathbf{Y}$ are simulation results for an equivalent design problem, it may be difficult to transfer knowledge from one dataset to the other since both sets of fields may not be represented in the same space, i.e., may have different dimensionality or topology. However, due to their intrinsic similarity, it is reasonable to assume that both $\mathbf{X}$ and $\mathbf{Y}$ share a common low-dimensional latent space in which the embedded data can be compared. The purpose of manifold alignment [25, 26] is to uncover such a shared latent space assuming it exists, thus enabling knowledge transfer between both datasets.

In the field of machine learning, manifold alignment has been applied to a variety of transfer learning problems such as protein alignment [32], image matching [25], magnetic resonance image classification [33], hyperspectral imaging visualization [34], etc. In their general framework, Wang and Mahadevan [26] define two types of manifold alignment approaches. The first type begins by identifying the intrinsic latent space for each of the individual datasets using any form of dimensionality reduction. Then, both latent spaces are aligned using a Procrustes analysis of the embedded data. For simplicity, we will refer to this approach as the Procrustes manifold alignment. In the second type, a shared latent space is uncovered in a semi-supervised manner by constructing a joint graph Laplacian. The resulting latent space attempts to preserve the individual and shared features of both datasets simultaneously.

An interesting feature of the first type of manifold alignment based on Procrustes analysis is that it can readily be combined with most conventional dimensionality reduction methods such as the POD. Therefore, this method lends itself to a relatively straightforward multi-fidelity extension of most POD-based ROMs, as it is later shown. This work focuses on the Procrustes manifold alignment-based approach and the investigation of semi-supervised techniques for multi-fidelity ROMs is deferred to future work.

1. Formulation

Let $\mathbf{X} \in \mathbb{R}^{d\times n}$ and $\mathbf{Y} \in \mathbb{R}^{q\times m}$ be matrices containing the results of a high- and low-fidelity model respectively, and with $n < m$. Additionally, let the n first solutions $\mathbf{y}_j$ and $\mathbf{x}_j$ have a pair-wise correspondence, i.e., where $\mathbf{y}_j$ and $\mathbf{x}_j$ are generated with the same design parameter values. The low-fidelity results are thus partitioned into $\mathbf{Y} = [\mathbf{Y}_L, \mathbf{Y}_U]$, where $\mathbf{Y}_L \in \mathbb{R}^{q\times n}$ contains the solutions linked to $\mathbf{X}$ while $\mathbf{Y}_U \in \mathbb{R}^{q\times(m-n)}$ holds the unlinked data.

The first step of the Procrustes manifold alignment is to find low-dimensional latent spaces (or manifolds) of the individual datasets. As a result, we define $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n] \in \mathbb{R}^{k\times n}$ and $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_m] \in \mathbb{R}^{k\times m}$ as the k-dimensional latent variables of $\mathbf{X}$ and $\mathbf{Y}$ respectively. As previously, we partition the low-fidelity data into $\mathbf{W} = [\mathbf{W}_L, \mathbf{W}_U]$ where $\mathbf{W}_L \in \mathbb{R}^{k\times n}$ and $\mathbf{W}_U \in \mathbb{R}^{k\times(m-n)}$ are the linked and unlinked latent variables. In practice, any dimensionality method could be used for this step, but since POD is the *de facto* method for most non-intrusive ROMs, it is the method that will be considered in the current work. The variables $\mathbf{Z}$ and $\mathbf{W}$ are, thus, given by

$$\mathbf{Z} = \mathbf{\Phi}_k^T \mathbf{X} \tag{7}$$

$$\mathbf{W} = \mathbf{\Psi}_k^T \mathbf{Y} \tag{8}$$

where $\mathbf{\Phi}_k \in \mathbb{R}^{d\times k}$ and $\mathbf{\Psi}_k \in \mathbb{R}^{q\times k}$ are orthonormal bases formed with the k-first POD modes of $\mathbf{X}$ and $\mathbf{Y}$ respectively.

Once both low-dimensional embeddings are identified, the following step is to find an affine transformation (i.e., translation, rotation, and scaling) of $\mathbf{W}$ that is optimally aligned with $\mathbf{Z}$. More specifically, let $\mathbf{Z}_{\text{lf}}$ denotes the transformed representation of $\mathbf{W}$, we seek the transformation which minimizes $\|\mathbf{Z} - \mathbf{Z}_{\text{lf},L}\|$. This becomes an orthogonal Procrustes problem, which is solved using the procedure given by Wang and Mahadevan [32]. For the sake of completeness, this process is repeated below with minor alterations to accommodate the current problem:

1) Translate the linked data $\mathbf{Z}$ and $\mathbf{W}_L$ such that their centroid is at the origin. Note, the mean of $\mathbf{Z}$ should already be zero since we assume $\mathbf{X}$ has been centered prior to the POD (see Section II.A). However, this is not necessarily true for $\mathbf{W}_L$ and we must shift $\mathbf{W}$ such that

$$\overline{\mathbf{w}}_L = \frac{1}{n}\sum_{j=1}^{n} \mathbf{w}_j = \mathbf{0} \tag{9}$$

2) Compute the Singular Value Decomposition (SVD) of $\mathbf{W}_L\mathbf{Z}^T$ such that

$$\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = \mathbf{W}_L\mathbf{Z}^T \tag{10}$$

where $\mathbf{U} \in \mathbb{R}^{k\times k}$ and $\mathbf{V} \in \mathbb{R}^{k\times k}$ are matrices containing the left and right singular vectors, while $\boldsymbol{\Sigma} \in \mathbb{R}^{k\times k}$ is a diagonal matrix containing the singular values.

3) Calculate $\mathbf{Z}_{\text{lf}}$ using the following mapping

$$\mathbf{Z}_{\text{lf}} = s\,\mathbf{Q}\mathbf{W} \tag{11}$$

where

$$s = \frac{\text{tr}(\boldsymbol{\Sigma})}{\text{tr}\left(\mathbf{W}_L\mathbf{W}_L^T\right)} \tag{12}$$

$$\mathbf{Q} = \mathbf{V}\mathbf{U}^T \tag{13}$$

Note that in the mapping of Eq. (11), $\mathbf{Q} \in \mathbb{R}^{k\times k}$ can be viewed as a rotation matrix while s is a scaling factor. In the end, manifold alignment yields a representation of the low-fidelity data $\mathbf{Y}$ in the latent space of the high-fidelity data $\mathbf{X}$. Since both datasets are represented in the same space, they can be easily compared and combined into a multi-fidelity model.

2. Datasets with Different Latent Dimensionality

In the previous section, we have assumed that the dimensionality of $\mathbf{Z}$ and $\mathbf{W}$ was identical. However, due to a larger amount of data, the dimensionality of $\mathbf{W}$ can conceivably be larger than $\mathbf{Z}$. The algorithm of Section II.B.1 can still be applied in such a case, thus making use of all the information available. The manifold alignment then becomes a projection Procrustes problem [35] such that $\mathbf{Q}$ is instead a projection matrix. In the opposite scenario where the dimensionality of $\mathbf{W}$ is less than $\mathbf{Z}$, the current algorithm can also be used by artificially extending the dimensionality of $\mathbf{W}$. This is achieved by padding the missing dimensions of $\mathbf{W}$ with zero vectors. Note that in such a case, the information transfer from the low- to the high-fidelity data will likely be weaker.

C. Multi-Fidelity Regression

With non-intrusive ROMs, new fields are predicted by estimating their coordinates in the latent space using some regression model. Assuming a POD-based approach, the formulation of a non-intrusive ROM is given by

$$\widetilde{\mathbf{x}} = \sum_{i=1}^{k} \boldsymbol{\phi}_i\, g_i(\mathbf{p}) \tag{14}$$

where $g_i : \mathbf{p} \mapsto \tilde{z}_i$ is a regression model that estimates the i-th coordinate $\tilde{z}_i$ in the POD latent space given some design parameters $\mathbf{p} \in \mathbb{R}^b$. In practice, g_i can be generated using most conventional surrogate modeling methods. Examples of non-intrusive ROMs using cubic splines [5], Radial-Basis Functions (RBF) [36, 37], Kriging [6, 38], and artificial neural-networks [7, 8] can be found in the literature. In the current work, single-fidelity ROM results are obtained using a Kriging model.

For a multi-fidelity ROM, the regression model g_i must also have a multi-fidelity formulation since we wish to leverage both high- and low-fidelity latent variables, i.e., combining both $\mathbf{Z}$ and $\mathbf{Z}_{\text{lf}}$. From the literature on multi-fidelity surrogate models, popular methods include bridge functions with either a multiplicative [39] or additive [40, 41] correction, and CoKriging [42–44]. For the proposed approach, the multi-fidelity regression of latent variables is realized using Hierarchical Kriging. As shown by Han and Görtz [45], Hierarchical Kriging offers an accuracy on par with CoKriging, but is relatively simpler to implement. The reader is referred to the work of Han and Görtz for details on the formulation of Hierarchical Kriging.

In should be noted that for all the experiments contained in this work, both the single- and multi-fidelity Kriging models are constructed using a Matérn 3/2 covariance function and are trained using the maximum likelihood criterion [46].

D. Proposed Multi-Fidelity ROM

The main contribution of this work is the formulation of a new multi-fidelity, parametric, and non-intrusive ROM based on manifold alignment. We refer to this new method as the Manifold-Alignment Reduced-Order Model

(MA-ROM). The overall procedure for the proposed method is described in Fig. 2. We denote the high- and low-fidelity models as $f_x : \mathbf{p} \mapsto \mathbf{x}$ and $f_y : \mathbf{p} \mapsto \mathbf{y}$ respectively, where the design parameters $\mathbf{p}$ are common to both models. The method is divided into two phases: an *offline* phase to train the model and an *online* phase to make new predictions. For the offline phase, the main steps are:

1) *Generate Linked Data:* Create n design vectors $\mathbf{P}_L = [\mathbf{p}_1, \ldots, \mathbf{p}_n] \in \mathbb{R}^{b \times n}$ and sample both $f_x(\mathbf{p})$ and $f_y(\mathbf{p})$ to generate the linked datasets $\mathbf{X}$ and $\mathbf{Y}_\mathrm{L}$.
2) *Generate Unlinked Data:* Create $m - n$ additional design vectors $\mathbf{P}_U = [\mathbf{p}_{n+1}, \ldots, \mathbf{p}_m] \in \mathbb{R}^{b \times (m-n)}$ and sample $f_y(\mathbf{p})$ only to generate the unlinked low-fidelity data $\mathbf{Y}_\mathrm{U}$.
3) *Compute the POD:* Given some desire RIC value, apply the POD to $\mathbf{X}$ and $\mathbf{Y}$ to obtain their respective POD modes, i.e., $\mathbf{\Phi}_k$ and $\mathbf{\Psi}_k$, as well as their latent variables $\mathbf{Z}$ and $\mathbf{W}$. Note, since we are only interested in predicting high-fidelity results, $\mathbf{\Psi}_k$ can be discarded.
4) *Align Manifolds:* Apply Procrustes manifold alignment to find the optimal transformation of $\mathbf{W}$ into $\mathbf{Z}_\mathrm{lf}$.
5) *Fit Regression Model:* Train the multi-fidelity regression models $\mathbf{g}(\mathbf{p}) = [g_1(\mathbf{p}), \ldots, g_k(\mathbf{p})]$ by combining both $\mathbf{Z}$ and $\mathbf{Z}_\mathrm{lf}$.

In the online phase, the process is identical to most non-intrusive ROMs and is described by the following steps:

6) *Predict Latent Variable:* Given an out-of-sample input $\mathbf{p}^*$, predict the corresponding latent variable $\tilde{\mathbf{z}}^*$ using $\mathbf{g}(\mathbf{p})$.
7) *Reconstruct Field:* Using $\mathbf{\Phi}_k$ and Eq. (7), compute the predicted value of the new high-fidelity solution $\tilde{\mathbf{x}}^*$ given $\tilde{\mathbf{z}}^*$.

In this study, the design samples in steps 1 and 2 are generated using Latin Hypercube Sampling (LHS), although any space-filling design of experiments could readily be used. Before proceeding, we highlight the salient features of the novel MA-ROM method:

- By reducing the dimensionality of $\mathbf{X}$ and $\mathbf{Y}$ separately and then aligning both subspaces, both sets of solutions do not need to be consistent in terms of dimensionality or topology. Therefore, the proposed method does not require any pre-processing and is agnostic to the representation of the individual fields. Although, the low- and high-fidelity fields must still be somewhat correlated for any multi-fidelity model to be effective.
- Since all the high- and low-fidelity solutions are projected onto the same latent space, the number of samples available for the training of the regression models $\mathbf{g}(\mathbf{p})$ can be relatively large. This will aid in reducing the prediction error of new latent space coordinates, especially for problems where the number of design parameters is large.

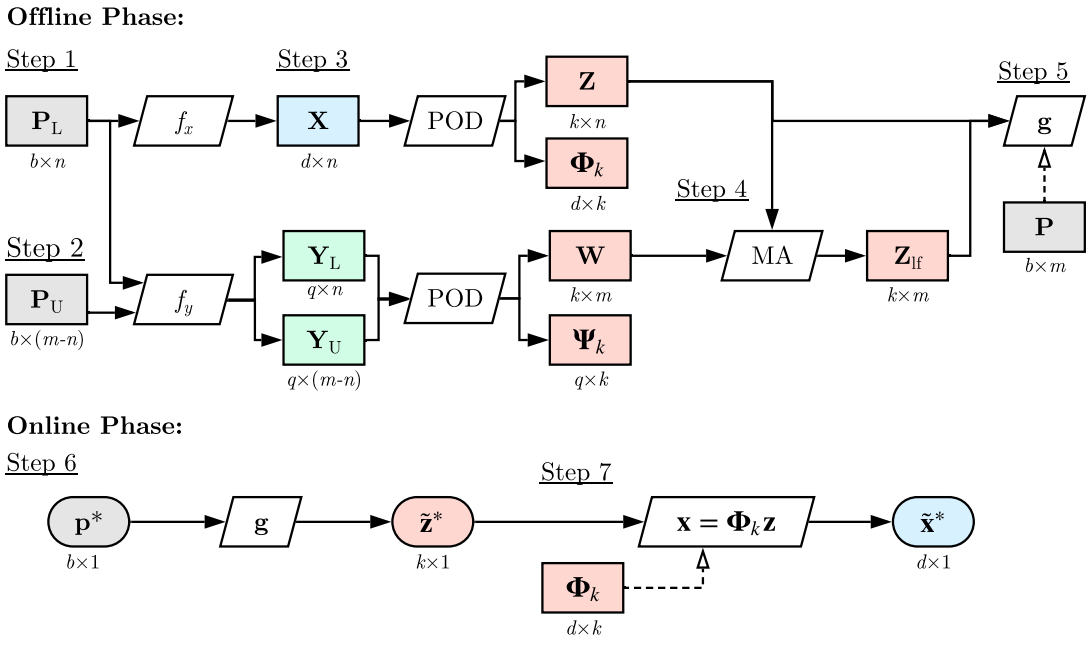


Fig. 2 Flowchart of the proposed MA-ROM method. This process is broken down into an offline (training) and online (prediction) phase.

- During manifold alignment, the POD modes $\mathbf{\Phi}_k$ of the high-fidelity data is left untouched, and as such, MA-ROM will not improve the accuracy of the dimensionality reduction in terms of reconstruction error (see Eq. (16)). Consequently, the proposed approach will mostly benefit situations where the prediction error of a non-intrusive ROMs is dominated by the accuracy of the underlying regression model (see Eq. (17)), which is often the case.

III. Test Case: Transonic Airfoil

To assess and demonstrate the benefits of MA-ROM, the proposed method is applied to the aerodynamic analysis of the RAE 2822 transonic airfoil [47]. In this experiment, the performance of MA-ROM is evaluated using different fidelity combinations, parametrizations, and sample sizes. The ROM shown in this study will focus on the prediction of the pressure coefficient C_P in the flow field surrounding the airfoil. A freestream Mach number of $M_\infty = 0.725$ is used for all of the simulations.

A. Simulation Setup

We use a compressible CFD simulation of the RAE 2822 to analyze the flow field around this airfoil in transonic conditions. The simulations are performed using the open-source SU2 code [48], and we solve for the turbulent flow-field using the Reynolds-Averaged Navier-Stokes (RANS) equations and the Spalart-Allmaras (SA) turbulence model [49]. The Jameson-Schmidt-Turkel (JST) scheme [50] is used for the spatial discretization and the steady-state solution is obtained using a backward Euler scheme. The fluid domain around the airfoil is discretized using a structured O-grid topology generated using a hyperbolic solver, an example of which is shown in Fig. 3. A grid convergence study is performed to evaluate the accuracy of the simulation and the results are reported in Table 1. These results are also compared to the experimental results of Cook et al. [47]. Note that the wind tunnel experiment of the RAE 2822 used an angle of attack of $\alpha = 2.92°$ whereas the current CFD simulation uses $\alpha = 2.5°$ to account for the walls interference [51]. The data in Table 1 shows that the CFD results asymptotically converge from the coarse to the fine grid. Also, the CFD values show good agreement with experimental results. For this study, the baseline grid (41,796 nodes) will be used as the high-fidelity model.

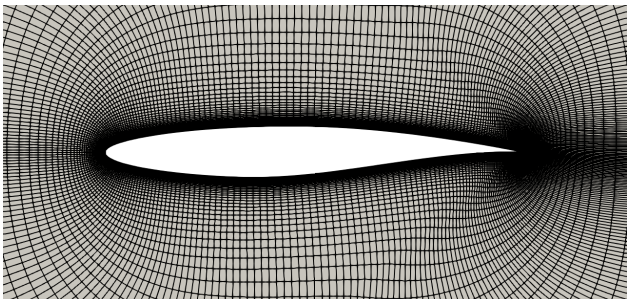


Fig. 3 Baseline O-grid of the RAE 2822 generated using a hyperbolic solver.

Table 1 Grid convergence study of the RAE 2822 CFD and comparison to experimental results [47]. The flow conditions correspond to Case 6 of the experimental campaign ($Re = 6.5 \times 10^6$ and $M_\infty = 0.725$).

Grid Level	Grid Size	α [deg]	C_n	C_m	C_d [10^4]
Experimental	*N/A*	*2.92*	*0.743*	*-0.095*	*127.0*
Fine	166,536	2.50	0.746	-0.085	122.5
Baseline	41,796	2.50	0.730	-0.082	122.8
Coarse	10,530	2.50	0.717	-0.081	132.6

B. Parametrization

The angle of attack and the airfoil shape are considered as the design parameters for this test case. The airfoil shape is parametrized using a Free Form Deformation (FFD) approach as illustrated in Fig. 4. This type of parametrization encloses a geometry into an FFD volume (or box), which is a rectangle in this test case. By displacing the boundary nodes of the FFD volume, the embedded geometry deforms and the number of boundary nodes defines the number of degrees of freedom of the parametrization. To prevent the FFD deformation from affecting the angle of attack of the airfoil, the nodes at the leading and trailing edge are kept fixed. We also limit the FFD control points to move along the vertical axis only. Once the new airfoil geometry is generated, the deformation is propagated into the rest of the grid using a pseudo-elastic approach. The tools required for this FFD parametrization are provided by the SU2 code.

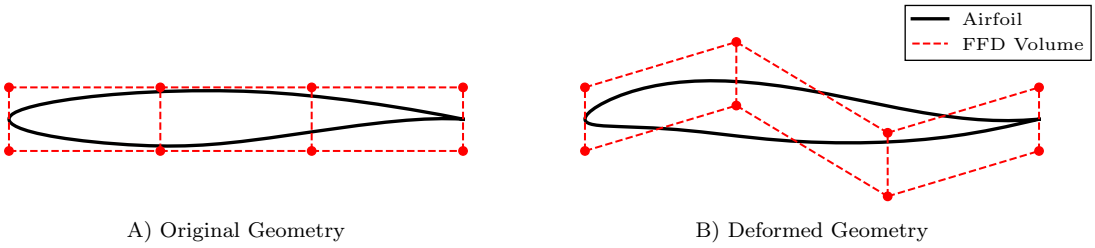


Fig. 4 Free form deformation of the RAE 2822 airfoil using a 4×2 FFD volume.

In this test case, two parametrization scenarios are considered. In addition to the angle of attack, the first parametrization uses two FFD control points (3×2 FFD box) while the second uses four FFD control points (4×2 FFD box, see Fig. 4). Therefore, the two parametrizations have an input space dimensionality of $b = 3$ and $b = 5$ respectively. In both cases, the angle of attack is varied between $0°$ and $4°$, while the FFD control points displacement is limited to ± 0.03 chord length. As a consequence of the curse of dimensionality, the higher-dimensional parametrization should require more samples to obtain an accurate ROM. A multi-fidelity ROM is expected to be especially beneficial in that case.

C. Error Metrics

The performance of the MA-ROM is quantified by comparing its prediction accuracy to a conventional POD-based ROM. Given n_{test} additional fields corresponding to out-of-sample designs, i.e., not used during the training of the model, we compute the normalized root-mean-squared (RMS) prediction error ϵ_{p} as follows

$$\epsilon_{\text{p}} = \sqrt{\frac{\sum_{j=1}^{n_{\text{test}}} \|\mathbf{x}_j^* - \tilde{\mathbf{x}}_j\|^2}{\sum_{j=1}^{n_{\text{test}}} \|\mathbf{x}_j^* - \bar{\mathbf{x}}\|^2}} \tag{15}$$

where $\mathbf{x}_j^*$ is the j-th out-of-sample solution and $\tilde{\mathbf{x}}_j$ is the corresponding ROM prediction. One should note that the normalization factor used in Eq. (15) is effectively dividing the RMS error by the total standard deviation of the testing dataset.

Recognizing that a non-intrusive ROM is essentially a combination between a dimensionality reduction model and a regression model, we decompose the prediction error into a reconstruction error and a regression error. The reconstruction error ϵ_{rc} is defined as the error linked to the mapping between the latent and physical space, or in the case of POD, the error associated with the accuracy of the POD modes. The normalized RMS reconstruction error ϵ_{rc} is given by

$$\epsilon_{\text{rc}} = \sqrt{\frac{\sum_{j=1}^{n_{\text{test}}} \|(\mathbf{I} - \boldsymbol{\Phi}_k \boldsymbol{\Phi}_k^T)\mathbf{x}_j^*\|^2}{\sum_{j=1}^{n_{\text{test}}} \|\mathbf{x}_j^* - \bar{\mathbf{x}}\|^2}} \tag{16}$$

Note that ϵ_{rc} is different from the reconstruction residual ε defined in Eq. (6) since the latter is computed from the training dataset. As for the regression error, it is characterized as the discrepancies in the latent space coordinates caused by the inaccuracies in the underlying regression models $\mathbf{g}(\mathbf{p})$. Therefore, we define the normalized RMS regression

error ϵ_{rg} as

$$\epsilon_{\text{rg}} = \sqrt{\frac{\sum_{j=1}^{n_{\text{test}}} \|\boldsymbol{\Phi}_k(\boldsymbol{\Phi}_k^T \mathbf{x}_j^* - \widetilde{\mathbf{z}}_j)\|^2}{\sum_{j=1}^{n_{\text{test}}} \|\mathbf{x}_j^* - \overline{\mathbf{x}}\|^2}} \tag{17}$$

where $\widetilde{\mathbf{z}}_j$ is the predicted latent space coordinates corresponding to $\mathbf{x}_j^*$. One can also view the reconstruction and regression errors as being the components of the prediction error that are respectively perpendicular and aligned with the POD subspace. As a result, the errors ϵ_{rc} and ϵ_{rg} can both be combined into ϵ_{p} as follows

$$\epsilon_{\text{p}} = \sqrt{\epsilon_{\text{rc}}^2 + \epsilon_{\text{rg}}^2} \tag{18}$$

D. Multi-Fidelity Datasets

This section lists the various fidelity levels that will be combined into multi-fidelity ROMs. The purpose of the presented fidelity levels is twofold: to demonstrate the computational savings achievable with a multi-fidelity ROM and to exhibit the capability of MA-ROM to fuse fields with disparate representations.

1. Fidelity Levels with Lower Computational Costs

The main purpose of a multi-fidelity model is to combine expensive simulation results with a cheaper auxiliary source of data to obtain an accurate model at a lower overall cost. As such, the present study will consider the following datasets of different fidelity levels and computational costs:

- *L1*: Flow field C_P using RANS CFD with the baseline grid (41,796 nodes);
- *L2*: Flow field C_P using RANS CFD with the coarse grid (10,530 nodes);
- *L3*: Flow field C_P using Euler CFD with a coarse grid and no near-wall refinement (8,910 nodes).

For all the ROMs used in this study, the fidelity L1 is used as the high-fidelity model and is combined with one of the other fidelity levels to construct a multi-fidelity model. Table 2 provides a comparison of the computational cost required to generate samples for each fidelity levels. It is worth noting that the execution times of the fidelities L2 and L3 are roughly 85% and 90% faster than L1. Therefore, great computational savings can potentially be achieved by leveraging these lower-fidelity options in a multi-fidelity ROM. Furthermore, the above datasets all use grids of different sizes, and thus, have inconsistent dimensionality. Other multi-fidelity ROM methods would require a pre-processing step where all the solutions would be mapped unto a common grid, usually the coarser one [18–20]. However, with the MA-ROM method, these results are combined into a single model without additional manipulations of the results.

2. Fidelity Levels with Inconsistent Representations

One of the key features of MA-ROM is that its formulation is agnostic to the representation of the low-fidelity results. In theory, any disparate sources of data can be fused with this method as long as they share some correlation. To highlight this feature, we define two additional datasets represented in truly different spaces:

- *L1-M*: Flow field Mach number M around the airfoil;
- *L1-S*: C_P distribution at the airfoil surface only.

Both of the above datasets use results from RANS CFD simulations with the baseline grid and are combined with the fidelity L1 to construct a multi-fidelity ROM of the flow field C_P distribution. Here, although L1-M and L1-S are not strictly speaking of lower-fidelity than L1, we assume that they represent data available in greater quantity and are correlated to L1 such that they can augment its prediction. One should note that for the dataset L1-M, even though the results use the same grid as L1, the physical quantities of the fields are different (i.e., C_P vs. M) and are not contained in the same latent space. As for the fidelity L1-S, the flow field C_P of L1 cannot be easily extrapolated from the results at the airfoil surface. With previous multi-fidelity methods, pre-processing the fields of L1-M and L1-S such that they can be combined with the results of L1 into a multi-fidelity ROM would be challenging. Despite the obvious inconsistencies in the fields, the MA-ROM method can fuse all of these datasets and exploit the underlying correlation, if it exists, without significant difficulties.

IV. Results and Discussion

This section presents and discusses the results of three distinct experiments using the RAE 2822 test case. The first experiment compares the computational cost and accuracy of MA-ROM to those of a single-fidelity ROM using

Table 2 Description of the various fidelity levels.

Fidelity Level	Physics Model	Field	Grid Size	Cost/Sample [CPU-min]*
L1	RANS	C_P	41,796	8.191
L2	RANS	C_P	10,530	1.201
L3	Inviscid	C_P	8,910	0.760
L1-M	RANS	M	41,796	-
L1-S	RANS	C_P	324†	-

*CPU times are based on an Intel Xeon Gold 6248 CPU
†Airfoil surface only

the $b = 3$ parametrization together with the fidelities L1, L2, and L3. The second experiment extends this comparison by also considering a higher dimensional parametrization with $b = 5$. Then, the third and last experiment demonstrates the capability of MA-ROM to fuse fields with inconsistent representation by combining the fidelity L1 to the fidelities L1-M and L1-S. A visual and qualitative comparison between the predicted fields of the three experiments is also presented at the end of this section.

To measure the model prediction accuracies for all the experiments of this study, a large dataset of 2,500 CFD simulations is compiled for each fidelity levels listed in Table 1 using an LHS design. A subset of n training cases is randomly selected from the L1 dataset and is used to train a conventional ROM model. For the training of an MA-ROM model, the same n cases and an additional $m - n$ cases are selected from one of the auxiliary datasets. The remaining cases from the L1 dataset are then used as a testing set to measure the prediction error for both the ROM and MA-ROM models. This process is repeated at least 50 times for each model configuration, and the average prediction error is recorded for each combination of training sample size and fidelity. For the single-fidelity ROM, values of n ranging from 10 to 1000 are evaluated. As for the multi-fidelity MA-ROM, the values of m are scaled proportionally with n via the multi-fidelity ratio $\tau = m/n$. Ratios of $\tau = \{2, 4, 8\}$ are considered together with the constraint $m \leq 1200$. Also, when applying POD to both the high- and low-fidelity datasets, the number of modes is selected by applying the criterion RIC $\geq 99.9999\%$ (see Eq. 5) for all the models.

A. Accuracy and Cost Trade-Off

The average prediction errors $\bar{\epsilon}_p$ for the single-fidelity ROM and the MA-ROM methods as a function of the number of high-fidelity training samples and the total training cost is shown in Fig. 5 and 6. The former figure shows the multi-fidelity results joining the fidelity levels L1 (baseline RANS CFD) and L2 (coarse RANS CFD), while the latter presents the results of combining the data of L1 and L3 (coarse inviscid CFD). In both figures, the errors presented are for the prediction of the C_P field around the RAE 2822 with the $b = 3$ parametrization.

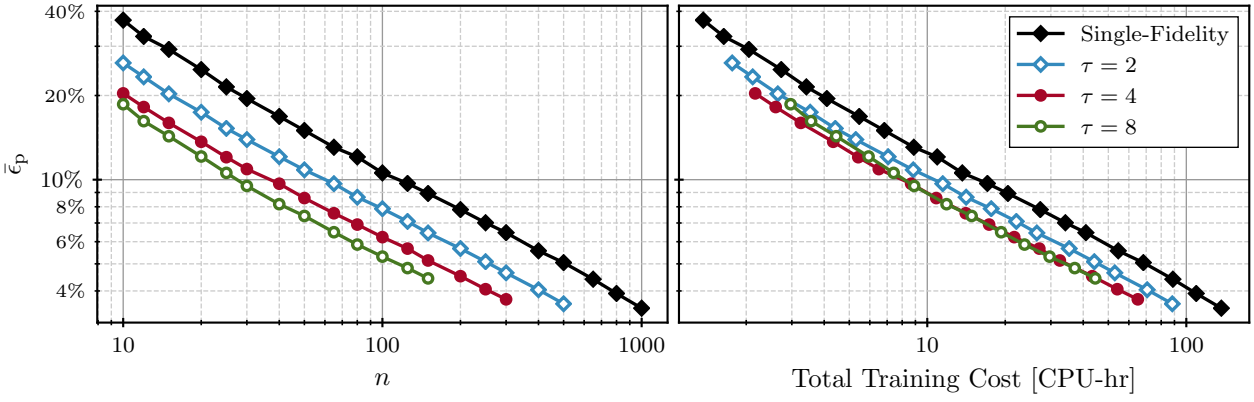


Fig. 5 Average prediction error $\bar{\epsilon}_p$ with respect to training cost for the multi-fidelity combination of the L1 and L2 datasets. Results are for the RAE 2822 test case with $b = 3$ and RIC = 99.9999%.

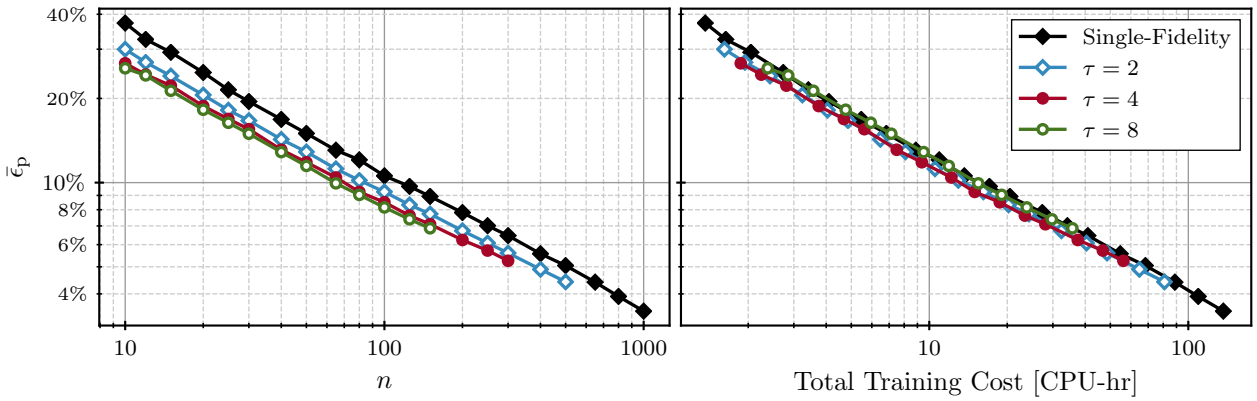


Fig. 6 Average prediction error $\bar{\epsilon}_p$ with respect to training cost for the multi-fidelity combination of the L1 and L3 datasets. Results are for the RAE 2822 test case with $b = 3$ and RIC = 99.9999%.

For all the considered models, Fig. 5 and 6 shows an exponential decrease in $\bar{\epsilon}_p$ as more training data is provided. We also observe that augmenting the high-fidelity data with either of the lower-fidelity datasets using MA-ROM yields a consistent reduction in $\bar{\epsilon}_p$ for a given n in all cases. We further note that increasing τ improves the accuracy of the multi-fidelity method, but it appears that the returns diminish as τ increases. Comparing Fig. 5 and 6 side by side, we observe that the overall performance of MA-ROM is higher when using the L2 rather than the L3 dataset. When compared to L2, the fidelity level L3 uses both a simpler physical model and a coarser grid for the aerodynamic analysis of the RAE 2822. Its correlation with the high-fidelity data is subsequently weaker, which in turn, hinders the multi-fidelity performance of MA-ROM. Nonetheless, the overall performance of the MA-ROM using both the L1 and L3 datasets remains better than a conventional ROM model using the L1 fidelity level alone.

To better understand the effects of the proposed multi-fidelity method, Table 3 offers a closer look at some of the results shown in Fig 5. For the current test case, we see from Table 3 that for a conventional ROM, the regression error $\bar{\epsilon}_{rg}$ is the dominant contributor to $\bar{\epsilon}_p$. Although high for small n, the reconstruction error $\bar{\epsilon}_{rc}$ is quickly reduced by increasing the training data size which allows for additional and more accurate POD modes. This is consistent with the high RIC criterion chosen for this study. As discussed in Sect. II.D, an MA-ROM trained with a comparable amount of high-fidelity data, say $n = 100$, will rely on an equivalent POD basis for the high-fidelity field reconstruction, and will accordingly have a similar $\bar{\epsilon}_{rc}$ than a single-fidelity model. However, the purpose of manifold alignment is to enrich the latent space with low-fidelity data to improve the prediction of the latent space coordinates. The results of Table 3 support this statement and indeed show a clear reduction in $\bar{\epsilon}_{rg}$ with additional low-fidelity data while $\bar{\epsilon}_{rc}$ remains more or less the same.

In practice, the availability of training data is limited by the number of computing resources at one's disposal. Fig. 5 and 6 also provide the prediction errors of the MA-ROM method as a function of the computational cost to generate the training data, taking into account the size of both the high- and low-fidelity datasets. The training cost is measured in terms of CPU time and uses the cost estimates from Table 2. With the L2 data as the low-fidelity results,

Table 3 Detailed breakdown of the prediction error of the ROM and MA-ROM method using L2 data. Results are for the RAE 2822 test case with $b = 3$ and RIC = 99.9999%.

Fidelity Levels	n	τ	m	$\bar{k}$	$\bar{\epsilon}_p$	$\bar{\epsilon}_{rc}$	$\bar{\epsilon}_{rg}$	Cost [CPU-hr]
L1	10	-	-	9	37.2%	16.0%	33.5%	1.37
	100			58	10.6%	1.03%	10.5%	13.7
	1000			92	3.47%	0.14%	3.47%	136.5
L1 + L2	100	2	200	58	7.87%	1.12%	7.79%	17.7
		4	400	58	6.23%	1.09%	6.13%	21.7
		8	800	58	5.31%	1.10%	5.19%	29.7

the results of Fig. 5 indicate that the MA-ROM method can provide a significant reduction in the training cost over a single-fidelity approach for a given $\bar{\epsilon}_p$. As for the data of Fig. 5, we once again observe that the L3 data offers more modest multi-fidelity improvement. In both cases, the cost of multi-fidelity models with $\tau = 8$ seems to be similar or worse than models with $\tau = 4$. This suggests there is a practical limit to the amount of information that the low-fidelity data can convey to the high-fidelity prediction.

To better establish the potential cost saving that MA-ROM can offer, Table 4 compares the total training cost of both single-fidelity ROMs and MA-ROMs for a target $\bar{\epsilon}_p$ value. These costs are estimated via a log-linear interpolation of the results of Fig. 5 and 6. The results of Table 4 shows that with $\tau = 4$, an MA-ROM fusing the L1 and L2 datasets requires 69% less high-fidelity samples, which translates into 57% less CPU time, than an equivalently accurate ROM using only L1 data. Using the lower-fidelity L3 in place of L2, the cost savings are cut down to 10% and 18% less CPU time for a target $\bar{\epsilon}_p$ respectively. Turning around the comparison, Table 5 presents the estimated $\bar{\epsilon}_p$ for a target cost. In this scenario, the data shows that for an equivalent training cost, the prediction of the C_P field around the RAE 2822 using a conventional ROM trained with the L1 data is 28% less accurate than an MA-ROM enhanced with the L2 dataset. Even with the L3 rather than the L2 data, the multi-fidelity ROM offers 5% to 9% better prediction given the same computational budget.

Table 4 Estimated cost for the ROM and MA-ROM method given a target prediction error. Results are for the RAE 2822 test case with $b = 3$ and RIC = 99.9999%.

Target ϵ_p	Fidelity Levels	b	n	τ	m	Cost [CPU-hr]
	L1		115	-	-	15.7
10%	L1 + L2	3	36 (−69%)	4	144	6.74 (−57%)
	L1 + L3		69 (−40%)	4	276	12.9 (−18%)
	L1		509	-	-	69.5
5%	L1 + L2	3	159 (−69%)	4	636	29.8 (−57%)
	L1 + L3		333* (−35%)	4	1,332*	62.3* (−10%)

*Extrapolated value

Table 5 Estimated prediction error for the ROM and MA-ROM method given a target computational budget. Results are for the RAE 2822 test case with $b = 3$ and RIC = 99.9999%.

Target Cost [CPU-hr]	Fidelity Levels	b	n	τ	m	$\bar{\epsilon}_p$
	L1		110	-	-	10.2%
15	L1 + L2	3	69	4	276	7.37% (−28%)
	L1 + L3		80	4	320	9.26% (− 9%)
	L1		440	-	-	5.34%
60	L1 + L2	3	277	4	1,108	3.87% (−28%)
	L1 + L3		321	4	1,284	5.09%* (− 5%)

*Extrapolated value

B. Effect of Input Space Dimensionality

The results in Fig. 7 show the accuracy and cost trade-off of the MA-ROM when applied to a design problem with 5 instead of 3 design parameters. For the sake of brevity, we only focus on the multi-fidelity model combining the L1 and L2 fidelity levels. Compared to Fig. 5, we observe an all-around increase in $\bar{\epsilon}_p$ and a slower decrease in the error with more training data for all the presented models. This degradation of both the ROM and MA-ROM performance is consistent with how empirical models generally behave with higher dimensional input spaces. That being said,

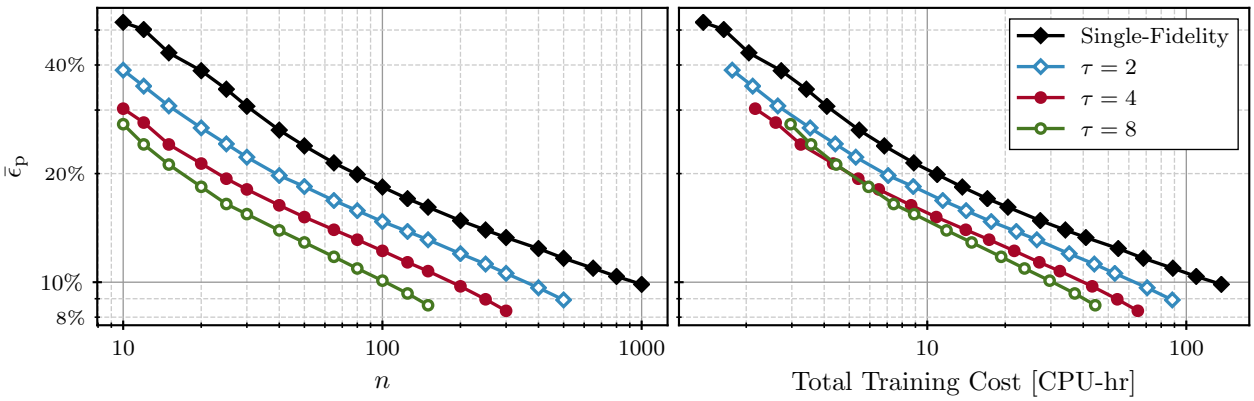


Fig. 7 Average prediction error $\bar{\epsilon}_p$ with respect to the training cost for the multi-fidelity combination of the L1 and L2 datasets. Results are for the RAE 2822 test case with $b = 5$ and RIC = 99.9999%.

the proposed multi-fidelity ROM method continues to provide a substantial reduction in the prediction error than an equivalent single-fidelity model. In fact, compared to the results of Fig. 5 where $b = 3$, the relative reduction in $\bar{\epsilon}_p$ in Fig. 7 with higher τ is in general equivalent or slightly improved.

For additional insights on the accuracy and cost trade-off of MA-ROM with a larger input space, Table 6 provides an estimate of the error and cost reductions that can be achieved for some target accuracy or computational budget. These estimates assume $\tau = 4$ and are obtained from the results of Fig. 7 via a simple log-linear interpolation. With a target $\bar{\epsilon}_p$ of 10%, Table 6 shows that augmenting the L1 fidelity level with the results of L2 can lower the number of high-fidelity samples by 80%, which results in a decrease of 73% of the total computational cost when compared to a single-fidelity model. In comparison, Table 4 presented a reduction of 57% for a similar scenario, but with 3 design parameters instead of 5. This greater cost saving for a fixed error is explained by the overall slower decrease in error with a higher n shown in Fig. 7. Fixing instead the computational budget to 60 CPU-hr, the results of Table 6 shows a potential decrease of 29% in $\bar{\epsilon}_p$, which is on par with the results of Table 5 for $b = 3$.

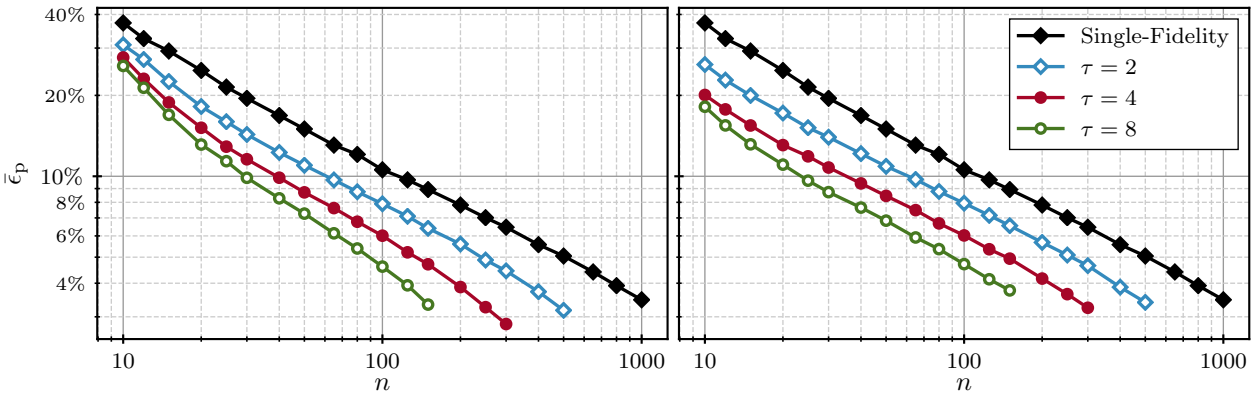
In summary, this experiment shows that the accuracy of MA-ROM is negatively affected by the addition of more design parameters, as most empirical models are due to the curse of dimensionality. Yet, the relative improvement that the multi-fidelity method offers over an equivalent single-fidelity ROM remains strong despite the additional parameters and can represent a substantial reduction in CPU time, especially if the high-fidelity model is computationally expensive.

Table 6 Estimated cost and prediction error for the ROM and MA-ROM method given either a target error or a computational budget. Results are for the RAE 2822 test case with $b = 5$ and RIC = 99.9999%.

Target ϵ_p	Fidelity Levels	b	n	τ	m	Cost [CPU-hr]
10%	L1	5	941	-	-	128.5
	L1 + L2		184 (−80%)	4	736	34.4 (−73%)
Target Cost [CPU-hr]	**Fidelity Levels**	b	n	τ	m	$\bar{\epsilon}_p$
60	L1	5	440	-	-	12.1%
	L1 + L2		277	4	1,108	8.60% (−29%)

C. Fields with Inconsistent Representation

While the previous experiments aimed at demonstrating the practical merits of MA-ROM as a multi-fidelity model, the purpose of this experiment is to exhibit the distinct capability of MA-ROM to fuse field data with different representations. As such, Fig. 8 presents the C_P prediction error obtained by combining the L1 fidelity level with either the L1-M (flow field Mach number M) or L1-S (Surface C_P distribution) datasets. A more detailed breakdown of the error for $n = 100$ is also provided in Table 7 for both multi-fidelity scenarios.



a) L1 + L1-M b) L1 + L1-S

Fig. 8 Average prediction error $\bar{\epsilon}_p$ with respect to n for the multi-fidelity combination of the L1 and either the L1-M or L1-S datasets. Results are for the RAE 2822 test case with $b = 3$ and RIC = 99.9999%.

Table 7 Detailed breakdown of the prediction error of the MA-ROM method using L1-M and L1-S data. Results are for the RAE 2822 test case with $b = 3$ and RIC = 99.9999%.

Fidelity Levels	n	τ	m	$\bar{k}$	$\bar{\epsilon}_p$	$\bar{\epsilon}_{rc}$	$\bar{\epsilon}_{rg}$
L1 + L1-M	100	2	200	58	7.90%	1.08%	7.82%
		4	400	58	6.01%	1.07%	5.91%
		8	800	58	4.61%	1.11%	4.47%
L1 + L1-S	100	2	200	58	7.94%	1.08%	7.86%
		4	400	58	6.02%	1.09%	5.92%
		8	800	58	4.71%	1.14%	4.57%

As with previous experiments, Fig. 8 shows that using either L1-M or L1-S as an auxiliary dataset to L1 still provides a noticeable reduction in $\bar{\epsilon}_p$ over the range of n and τ considered. We also note that the multi-fidelity benefits of MA-ROM remain higher with larger τ, although in the current scenarios, the improvement appears to be more modest at lower n, especially for the models using the L1-M dataset. Comparing the results of Table 7 to those of Table 3, we observe that using the L1-M or L1-S dataset instead of the L2 data provides similar, yet slightly lower, regression errors while leaving the reconstruction error unaffected. The marginal improvement in the results of Table 7 is not entirely surprising since the L1-M and L1-S datasets should have a relatively strong correlation with the L1 fidelity level. All three of those datasets are indeed produced by the same CFD setup, although they differ in terms of physical quantity and geometry of interest. However, the main outcome from this experiment is that the multi-fidelity performance of MA-ROM was not handicapped whatsoever by the inconsistencies existing between the training datasets. This confirms that the proposed MA-ROM method is viable even in situations where the high- and low-fidelity fields have different and incompatible representations, as long as some correlation exists between them.

D. Comparison of the Multi-Fidelity Predictions

To complement the results of the previous experiments, Fig. 9 offers a visual comparison between the fields predicted using the MA-ROM method with the various multi-fidelity combinations considered in this study. These fields are the predicted pressure coefficient $\tilde{C}_P$ generated using the L1 data as the high-fidelity dataset and either the L2, L3, L1-M, or L1-S data as the auxiliary dataset. A value of $n = 100$ and $\tau = 4$ was chosen for this qualitative visual comparison. These results are also compared to the prediction of an equivalent single-fidelity ROM and the actual C_P field. The discrepancies between the prediction and the actual results, i.e., between $\tilde{C}_P$ and C_P, are shown in Fig. 9 as well.

In the previous sections of this work, we noted that the MA-ROM method offered a lower prediction error than a conventional ROM for all the of multi-fidelity combinations given the same n. This observation seems to be supported

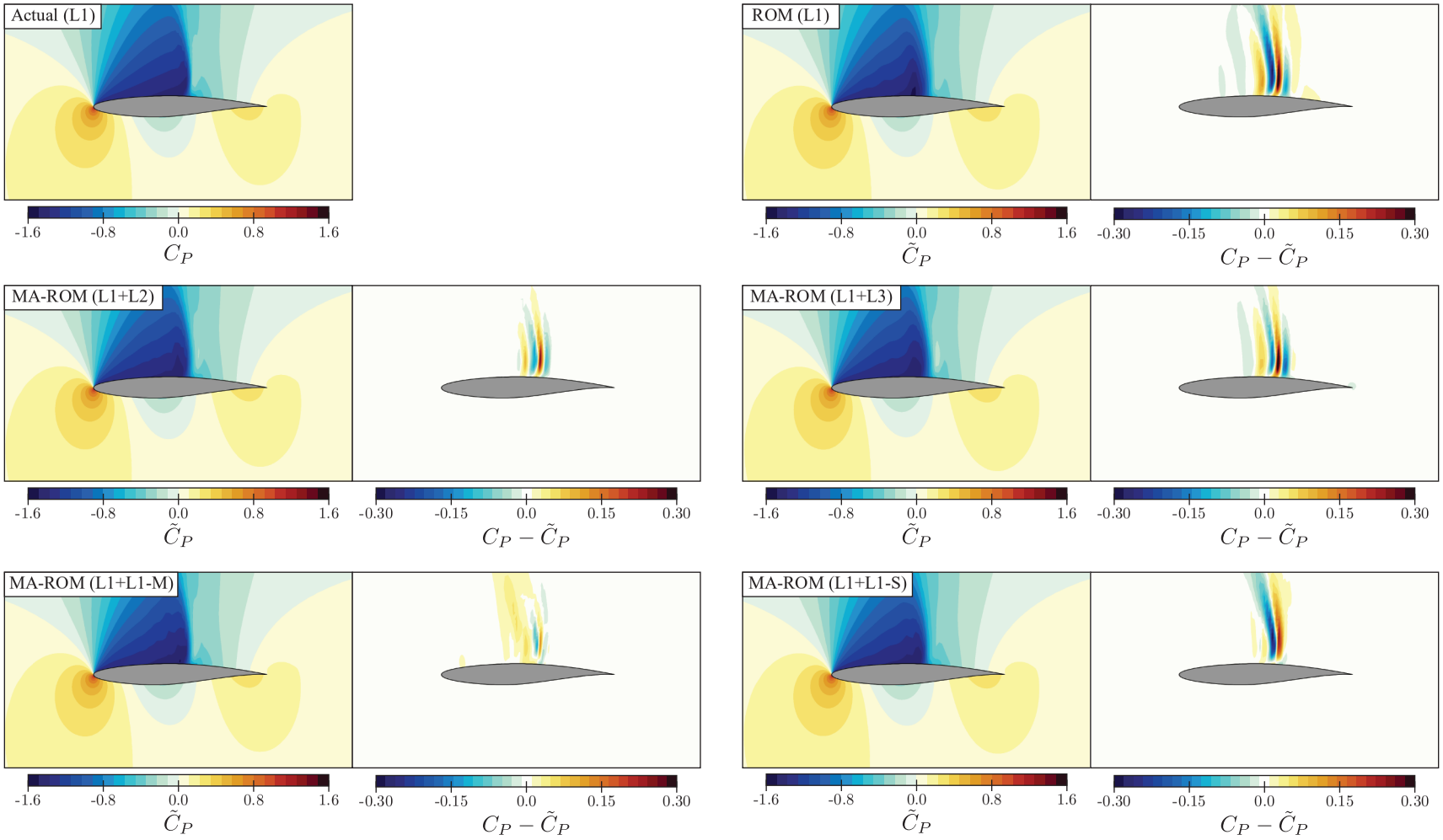


Fig. 9 Flow field visualization of the predicted C_P and the prediction discrepancy for the various multi-fidelity combinations of this study. Results are for the RAE 2822 test case with b = 3 and RIC = 99.9999%**.**

by Fig. 9 where all the presented MA-ROM predictions are qualitatively closer to the actual C_P distribution than a conventional ROM in a similar scenario. In addition, the MA-ROMs using the L2, L1-M, and L1-S fidelities as the auxiliary dataset appear to provide better predictions than the MA-ROM using the L3 data, which is consistent with the discussion of Sec. IV.A and IV.C. A glance at the error fields of Fig. 9 reveals that most of the prediction error in all the scenarios is concentrated near a strong shock on the airfoil upper surface. It is a known fact that POD-based ROMs tend to struggle with the prediction of discontinuities such as a shock wave [10]. In spite of this limitation, predictions from the proposed MA-ROM method result in an error field that is narrower and of lower intensity near this discontinuity, thus capturing the shock more accurately. All in all, the fields of Fig. 9 provide a visual confirmation of most of the observations made in section IV.

V. Conclusion

This work introduced a novel non-intrusive and multi-fidelity ROM method based on manifold alignment. Unlike previous multi-fidelity approaches, the developed MA-ROM method is capable of combining fidelity levels with disparate field representations. We assessed the performance of the MA-ROM method by applying it to a transonic airfoil problem and we observed reduction in the computational training cost between 10% and 73% when compared to a single-fidelity approach with comparable accuracy. Notwithstanding the reduced performance of any ROMs with larger design spaces, it was established that the multi-fidelity benefits of MA-ROM can persist with higher dimensional parametrizations. Finally, we demonstrated the flexibility of the method to handle inconsistent fields by successfully combining correlated fields representing different physical quantities and topologies.

Although this study mostly focused on aerodynamic predictions, the MA-ROM is also relevant to other disciplines having major topological differences between fidelity levels. For instance, with structural modeling, the solutions of beam, shell, and solid models each have distinct representation, but could potentially be combined into a single prediction model via MA-ROM. Furthermore, the current work only considered a POD-based approach, but manifold alignment can also serve as a multi-fidelity extension to other non-intrusive ROM methods. A potential candidate would be the non-linear isomap-based approach developed by Franz et al. [36], which was shown to offer superior predictions in the presence of discontinuities. Future research involving the MA-ROM method includes the above considerations.

References

[1] Lucia, D. J., Beran, P. S., and Silva, W. A., "Reduced-order modeling: New approaches for computational physics," *Progress in Aerospace Sciences*, Vol. 40, No. 1-2, 2004, pp. 51–117. https://doi.org/10.1016/j.paerosci.2003.12.001.

[2] Benner, P., Gugercin, S., and Willcox, K., "A survey of model reduction methods for parametric systems," *MPI Magdeburg Preprints*, Vol. MPIMD, No. 13-14, 2013, pp. 1–36. https://doi.org/10.1137/130932715, URL http://www2.mpi-magdeburg.mpg.de/preprints/2013/MPIMD13-14.pdf.

[3] Lumley, J. L., "The Structure of Inhomogeneous Turbulent Flows," *Atmospheric turbulence and radio propagation*, edited by A. M. Yaglom and V. I. Tatarski, Nauka, Moscow, 1967, pp. 166–178.

[4] Jolliffe, I., *Principal Component Analysis*, 2nd ed., Springer Series in Statistics, Springer-Verlag, New York, 2002. https://doi.org/10.1007/b98835, URL http://link.springer.com/10.1007/b98835.

[5] Bui-Thanh, T., Damodaran, M., and Willcox, K., "Proper Orthogonal Decomposition Extensions for Parametric Applications in Compressible Aerodynamics," *21st AIAA Applied Aerodynamics Conference*, , No. June, 2003, pp. 1–11. https://doi.org/10.2514/6.2003-4213, URL http://arc.aiaa.org/doi/10.2514/6.2003-4213.

[6] Fossati, M., "Evaluation of Aerodynamic Loads via Reduced-Order Methodology," *AIAA Journal*, Vol. 53, No. 8, 2015, pp. 2389–2405. https://doi.org/10.2514/1.J053755, URL http://arc.aiaa.org/doi/10.2514/1.J053755.

[7] Park, K. H., Jun, S. O., Baek, S. M., Cho, M. H., Yee, K. J., and Lee, D. H., "Reduced-Order Model with an Artificial Neural Network for Aerostructural Design Optimization," *Journal of Aircraft*, Vol. 50, No. 4, 2013, pp. 1106–1116. https://doi.org/10.2514/1.C032062.

[8] Swischuk, R., Mainini, L., Peherstorfer, B., and Willcox, K., "Projection-based model reduction: Formulations for physics-based machine learning," *Computers & Fluids*, Vol. 179, 2019, pp. 704–717. https://doi.org/10.1016/j.compfluid.2018.07.021, URL https://doi.org/10.1016/j.compfluid.2018.07.021.

[9] Rajaram, D., Puranik, T. G., Perron, C., and Mavris, D. N., "Non-Intrusive Parametric Reduced Order Modeling using Randomized Algorithms," *AIAA Scitech 2020 Forum*, 2020, p. 0417. https://doi.org/10.2514/6.2020-0417, URL https://arc.aiaa.org/doi/abs/10.2514/6.2020-0417.

[10] Li, J., and Zhang, W., "The performance of proper orthogonal decomposition in discontinuous flows," *Theoretical and Applied Mechanics Letters*, Vol. 6, No. 5, 2016, pp. 236–243. https://doi.org/10.1016/j.taml.2016.08.008, URL http://dx.doi.org/10.1016/j.taml.2016.08.008.

[11] Forrester, A. I., Sóbester, A., and Keane, A. J., "Multi-fidelity optimization via surrogate modelling," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 463, No. 2088, 2007, pp. 3251–3269. https://doi.org/10.1098/rspa.2007.1900.

[12] Toal, D. J. J., "Some considerations regarding the use of multi-fidelity Kriging in the construction of surrogate models," *Structural and Multidisciplinary Optimization*, Vol. 51, No. 6, 2015, pp. 1223–1245. https://doi.org/10.1007/s00158-014-1209-5, URL http://link.springer.com/10.1007/s00158-014-1209-5.

[13] Peherstorfer, B., Willcox, K., and Gunzburger, M., "Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization," *SIAM Review*, Vol. 60, No. 3, 2018, pp. 550–591. https://doi.org/10.1137/16M1082469, URL http://arxiv.org/abs/1806.10761https://epubs.siam.org/doi/10.1137/16M1082469.

[14] Yondo, R., Andrés, E., and Valero, E., "A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses," *Progress in Aerospace Sciences*, Vol. 96, No. November 2017, 2018, pp. 23–61. https://doi.org/10.1016/j.paerosci.2017.11.003.

[15] Mifsud, M., "Reduced-order modelling for high-speed aerial weapon aerodynamics," Ph.d. thesis, Cranfield University, 2008. URL http://hdl.handle.net/1826/3511.

[16] Mifsud, M. J., MacManus, D. G., and Shaw, S. T., "A variable-fidelity aerodynamic model using proper orthogonal decomposition," *International Journal for Numerical Methods in Fluids*, Vol. 82, No. 10, 2016, pp. 646–663. https://doi.org/10.1002/fld.4234.

[17] Bertram, A., Othmer, C., and Zimmermann, R., "Towards Real-time Vehicle Aerodynamic Design via Multi-fidelity Data-driven Reduced Order Modeling," *2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, , No. January, 2018. https://doi.org/10.2514/6.2018-0916, URL https://arc.aiaa.org/doi/10.2514/6.2018-0916.

[18] Benamara, T., Breitkopf, P., Lepot, I., and Sainvitu, C., "Multi-Fidelity Extension To Non-Intrusive Proper Orthogonal Decomposition Based Surrogates," *Proceedings of the VII European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS Congress 2016)*, , No. January 2016, 2016, pp. 4129–4145. https://doi.org/10.7712/100016.2098.9174, URL http://www.eccomasproceedia.org/conferences/eccomas-congresses/eccomas-congress-2016/2098.

[19] Benamara, T., Breitkopf, P., Lepot, I., Sainvitu, C., and Villon, P., "Multi-fidelity POD surrogate-assisted optimization: Concept and aero-design study," *Structural and Multidisciplinary Optimization*, Vol. 56, No. 6, 2017, pp. 1387–1412. https://doi.org/10.1007/s00158-017-1730-4.

[20] Malouin, B., Trépanier, J.-Y., and Gariépy, M., "Interpolation of Transonic Flows Using a Proper Orthogonal Decomposition Method," *International Journal of Aerospace Engineering*, Vol. 2013, 2013, pp. 1–11. https://doi.org/10.1155/2013/928904, URL http://www.hindawi.com/journals/ijae/2013/928904/.

[21] Toal, D. J. J., "On the potential of a multi-fidelity G-pod based approach for optimization & uncertainty quantification," *Turbine technical conference and exposition*, 2014.

[22] Benamara, T., Breitkopf, P., Lepot, I., and Sainvitu, C., "Adaptive infill sampling criterion for multi-fidelity optimization based on Gappy-POD," *Structural and Multidisciplinary Optimization*, Vol. 54, No. 4, 2016, pp. 843–855. https://doi.org/10.1007/s00158-016-1440-3, URL http://dx.doi.org/10.1007/s00158-016-1440-3http://link.springer.com/10.1007/s00158-016-1440-3.

[23] Everson, R., and Sirovich, L., "Karhunen–Loève procedure for gappy data," *Journal of the Optical Society of America A*, Vol. 12, No. 8, 1995, p. 1657. https://doi.org/10.1364/JOSAA.12.001657, URL https://www.osapublishing.org/abstract.cfm?URI=josaa-12-8-1657.

[24] Bui-Thanh, T., Damodaran, M., and Willcox, K. E., "Aerodynamic Data Reconstruction and Inverse Design Using Proper Orthogonal Decomposition," *AIAA Journal*, Vol. 42, No. 8, 2004, pp. 1505–1516. https://doi.org/10.2514/1.2159, URL http://arc.aiaa.org/doi/10.2514/1.2159.

[25] Ham, J., Lee, D. D., and Saul, L. K., "Semisupervised alignment of manifolds," *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, Vol. 10, 2005, pp. 120–127. URL http://www.cis.upenn.edu/{~}lsaul/abstracts.html{#}semi{_}aistats05.

[26] Wang, C., and Mahadevan, S., "A general framework for manifold alignment," *AAAI Fall Symposium Series*, 2009, pp. 53–58.

[27] Sirovich, L., "Turbulence and the dynamics of coherent structures. II. Symmetries and transformations," *Quarterly of Applied Mathematics*, Vol. 45, No. 3, 1987, pp. 573–582. https://doi.org/10.1090/qam/910463, URL http://www.ams.org/qam/1987-45-03/S0033-569X-1987-0910463-9/.

[28] Magnus, B., Landelius, T., and Knutsson, H., *A Unified Approach to PCA , PLS , MLR and CCA*, Linköping University, Department of Electrical Engineering, 1997.

[29] Pinnau, R., "Model Reduction via Proper Orthogonal Decomposition," *Model Order Reduction: Theory, Research Aspects and Applications*, Springer, 2008, pp. 95–109. https://doi.org/10.1007/978-3-540-78841-6_5, URL http://link.springer.com/10.1007/978-3-540-78841-6{_}5.

[30] Vendl, A., and Faßbender, H., "Projection-Based Model Order Reduction for Steady Aerodynamics," *Computational Flight Testing: Results of the Closing Symposium of the German Research Initiative ComFliTe, Braunschweig, Germany, June 11th-12th, 2012*, edited by N. Kroll, R. Radespiel, J. W. Burg, and K. Sørensen, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 151–166. https://doi.org/10.1007/978-3-642-38877-4_11, URL https://doi.org/10.1007/978-3-642-38877-4{_}11.

[31] Bishop, C. M., *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, Berlin, Heidelberg, 2006.

[32] Wang, C., and Mahadevan, S., "Manifold alignment using Procrustes analysis," *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 1120–1127. https://doi.org/10.1145/1390156.1390297, URL http://www-all.cs.umass.edu/{~}chwang/papers/ICML-2008.pdf.

[33] Guerrero, R., Ledig, C., and Rueckert, D., "Manifold Alignment and Transfer Learning for Classification of Alzheimer's Disease," 2014, pp. 77–84. https://doi.org/10.1007/978-3-319-10581-9_10, URL http://link.springer.com/10.1007/978-3-319-10581-9{_}10.

[34] Liao, D., Qian, Y., and Zhou, J., "Visualization of hyperspectral imaging data based on manifold alignment," *Proceedings - International Conference on Pattern Recognition*, 2014, pp. 70–75. https://doi.org/10.1109/ICPR.2014.22.

[35] Gower, J. C., and Dijksterhuis, G. B., *Procrustes Problems*, Oxford University Press, 2004. https://doi.org/10.1093/acprof:oso/9780198510581.001.0001, URL http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198510581.001.0001/acprof-9780198510581.

[36] Franz, T., Zimmermann, R., Görtz, S., and Karcher, N., "Interpolation-based reduced-order modelling for steady transonic flows via manifold learning," *International Journal of Computational Fluid Dynamics*, Vol. 28, No. 3-4, 2014, pp. 106–121. https://doi.org/10.1080/10618562.2014.918695.

[37] Kato, H., and Funazaki, K.-i., "POD-Driven Adaptive Sampling for Efficient Surrogate Modeling and its Application to Supersonic Turbine Optimization," *Volume 2B: Turbomachinery*, ASME, 2014, p. V02BT45A023. https://doi.org/10.1115/GT2014-27229, URL http://proceedings.asmedigitalcollection.asme.org/proceeding.aspx?doi=10.1115/GT2014-27229.

[38] Zimmermann, R., "Gradient-enhanced surrogate modeling based on proper orthogonal decomposition," *Journal of Computational and Applied Mathematics*, Vol. 237, No. 1, 2013, pp. 403–418. https://doi.org/10.1016/j.cam.2012.06.010, URL http://dx.doi.org/10.1016/j.cam.2012.06.010.

[39] Chang, K. J., Haftka, R. T., Giles, G. L., and Kao, I.-J., "Sensitivity-based scaling for approximating structural response," *Journal of Aircraft*, Vol. 30, No. 2, 1993, pp. 283–288. https://doi.org/10.2514/3.48278, URL http://arc.aiaa.org/doi/10.2514/3.48278.

[40] Choi, S., Alonso, J. J., and Kroo, I. M., "Two-Level Multifidelity Design Optimization Studies for Supersonic Jets," *Journal of Aircraft*, Vol. 46, No. 3, 2009, pp. 776–790. https://doi.org/10.2514/1.34362, URL http://arc.aiaa.org/doi/10.2514/1.34362.

[41] Tang, C., Gee, K., and Lawrence, S., "Generation of Aerodynamic Data using a Design Of Experiment and Data Fusion Approach," *43rd AIAA Aerospace Sciences Meeting and Exhibit*, , No. January, 2005, pp. 1–9. https://doi.org/10.2514/6.2005-1137, URL http://arc.aiaa.org/doi/10.2514/6.2005-1137.

[42] Kennedy, M., and O'Hagan, A., "Predicting the output from a complex computer code when fast approximations are available," *Biometrika*, Vol. 87, No. 1, 2000, pp. 1–13. https://doi.org/10.1093/biomet/87.1.1, URL papers2://publication/uuid/1A69EB40-F7E9-48E1-B609-FF3A7497DBC2http://biomet.oupjournals.org/cgi/doi/10.1093/biomet/87.1.1.

[43] Laurenceau, J., and Sagaut, P., "Building Efficient Response Surfaces of Aerodynamic Functions with Kriging and Cokriging," *AIAA Journal*, Vol. 46, No. 2, 2008, pp. 498–507. https://doi.org/10.2514/1.32308, URL http://arc.aiaa.org/doi/10.2514/1.32308.

[44] Han, Z., Zimmerman, R., and Görtz, S., "Alternative Cokriging Method for Variable-Fidelity Surrogate Modeling," *AIAA Journal*, Vol. 50, No. 5, 2012, pp. 1205–1210. https://doi.org/10.2514/1.J051243, URL http://arc.aiaa.org/doi/10.2514/1.J051243.

[45] Han, Z.-H., and Görtz, S., "Hierarchical Kriging Model for Variable-Fidelity Surrogate Modeling," *AIAA Journal*, Vol. 50, No. 9, 2012, pp. 1885–1896. https://doi.org/10.2514/1.J051354, URL http://arc.aiaa.org/doi/10.2514/1.J051354.

[46] Rasmussen, C. E., and Williams, C. K. I., *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, UNITED STATES, 2005. URL http://ebookcentral.proquest.com/lib/gatech/detail.action?docID=3338604.

[47] Cook, P., Mcdonald, M., and Firmin, M., "Experimental data base for computer program assessment," *AGARD AR-138*, 1979.

[48] Economon, T. D., Palacios, F., Copeland, S. R., Lukaczyk, T. W., and Alonso, J. J., "SU2: An Open-Source Suite for Multiphysics Simulation and Design," *AIAA Journal*, Vol. 54, No. 3, 2015, pp. 1–19. https://doi.org/10.2514/1.J053813, URL http://dx.doi.org/10.2514/1.J053813.

[49] Spalart, P., and Allmaras, S., "A one-equation turbulence model for aerodynamic flows," *30th Aerospace Sciences Meeting and Exhibit*, American Institute of Aeronautics and Astronautics, Reston, Virigina, 1992. https://doi.org/10.2514/6.1992-439, URL http://arc.aiaa.org/doi/10.2514/6.1992-439.

[50] Jameson, A., Schmidt, W., and Turkel, E., "Numerical solution of the Euler equations by finite volume methods using Runge Kutta time stepping schemes," *14th Fluid and Plasma Dynamics Conference*, American Institute of Aeronautics and Astronautics, Reston, Virigina, 1981. https://doi.org/10.2514/6.1981-1259, URL http://arc.aiaa.org/doi/10.2514/6.1981-1259.

[51] Garbaruk, A., Shur, M., Strelets, M., and Spalart, P. R., "Numerical study of wind-tunnel walls effects on transonic airfoil flow," *AIAA Journal*, Vol. 41, No. 6, 2003, pp. 1046–1054. https://doi.org/10.2514/2.2071.